DOCUMENT RESUME

ED 324 912 FL 018 872

AUTHOR Subramaniam, R.; Hock, Tan Soon
TITLE Towards Designing a Test of Communicative Performance--A Case Study.
PUB DATE 90
NOTE 28p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Case Studies; *Communicative Competence (Languages); *English (Second Language); *English for Special Purposes; Foreign Countries; Higher Education; *Language Tests; *Test Construction; Testing
IDENTIFIERS *University of Malaya (Malaysia)

ABSTRACT

The development of a communicative language competence test for the University of Malaya's Spoken English Project, which includes courses in basic English, English for Pre-Occupational Purposes, and Public Administration and Commercial English, is described. In the development process, a concerted effort was made to ensure that course goals and aims were reflected in the testing procedures. The situations presented in the oral interaction segments are simulations of target events in the professional world or real-life settings. The tasks set for the student demand the use of the language. The specification of criterion and target levels of performance are derived from course content, student data, target occupation operations, and other assessment scales. The criteria on which performance is assessed reflect the blend of communicative competence and performance in the materials. Improvement of test reliability has focused on refinement of the instrument, examiner training, and the training of "plants" in the testing situation, used to put students at ease and introduce unpredictability into interactions. Problems of test administration have not yet been resolved. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED324912

'TOWARDS DESIGNING A TEST OF COMMUNICATIVE PERFORMANCE - A CASE STUDY'

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
T. Hock
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

R. SUBRAMANIAM
and
TAN SOON HOCK

1. The Problem Area

The University of Malaya Spoken English Project launched in 1980 was designed specifically to answer the demands of the professional world for proficiency in speaking required of its present and future graduate employees. Its broad brief was to establish "a language teaching/learning programme .... that would lead to the design and production of a course that will provide selected students with a facility to improve their spoken English". The related problem was warded thus : "The problem currently noticed is that a significant proportion of our (University of Malaya) graduates is not as articulate in English as is considered necessary for them to be fully effective in their respective jobs". Data gathered from professionals and employers attest to this observation.

The problem of oral communication in English is acute. By 1983 the entire secondary school population

FL 018872

would have been taught through the medium of Bahasa Malaysia, the national language. By 1986 it is expected that undergraduate programmes at the University would be fully conducted through the national language. At best the only contact with English would be through the 4-6 hour per week of language classes, and the voluntary reference reading of English books. Yet, the graduate is, all too suddenly, expected to be able to communicate effectively in English when he assumes an appointment in the public or private sector. A profile of communication needs of, for example, a business executive, would include using English for communication in the main cities of Malaysia, and other ASEAN countries, Anglophone countries, on travels, etc. In Malaysia itself he might need to use it in commercial premises, banks and finance houses, government offices and social settings. The demands, therefore, on him are not commensurate with the entry level of proficiency he displays at the University, even after 11 years of studying English as a foreign language. The project therefore attempts to bridge this gap, to enable the students, with their differing levels of language competence, to master enough confidence and poise to <u>use</u> the kind of language they would encounter in real life in their respective careers.

Preliminary data gathering revealed that since a considerable number of undergraduates would still be uncertain as to what profession to pursue (especially

those from the Arts and Social Sciences) it would be necessary to plot a course that would not require too technical/professional a situation with its accompanying language use. Instead it would aim at providing a lighter professional content, emphasising more on confidence-building and activation of language skills so that the student would be equipped for entry into the profession and meeting initial training needs. Students will move from the Pre-Occupational Purposes (course to the Public Administration and Commercial English course (PACE).

The PACE course is designed to provide training in spoken English for general managerial purposes and is seen as in-service training for junior executives. Students will move from the POP course to PACE and graduates will be judged able to enter the PACE course only if they have completed the POP course or pass a PACE entry level test.

If syllabus design, course content and pedagogy places a priority on language use in particular real-life situations, different from the traditional emphasis on language form in discrete-point testing, then it makes sense that a learner should be tested in his ability to display actual production and comprehension of utterances. To do this he would need to be able to manipulate his knowledge of the language (competence) and to translate this into appropriate use

in a specified sociolinguistic/professional/academic setting. The advantage of this as a positive backwash of testing on the teaching programme cannot be understated, especially when the idea of learning English has a low appeal. However, it has its attendant problems. Canale (1981) expresses this point in the following manner "Evaluation within a communicative approach must address new means such as socio-linguistic appropriateness rules, new assessment formats to permit and encourage creative, open-ended language use, new test administration procedures to emphasize interpersonal interaction in authentic situations, and new scoring procedures of a more manual and judgemental nature".

This paper will describe how the project attempts to take up these challenges and demands in its direct-performance testing.

2. Rationale for course design

Generally all three courses in the University of Malaya Spoken English Project are organised on the basis of providing firstly, controlled opportunities for language use and secondly, extending and improving the quality of resources available for use. 'Controlled opportunities for language use' is defined in the following manner :

a. Opportunities for use should be controlled in terms of what the language is used for, not what language is used.

b. The opportunities for use should reflect target professional uses and reflect the range of uses to be encountered.

c. Opportunities for use cannot be defined in terms of isolated functions but as complete interactions involving the minimum roles of two participants and an outcome which enables the interaction to be evaluated.

'Improving the quality of language resources' refers to the practise of language features. These are not practised for their own sake but as resources for performance. This support component is aimed at

: extending and improving the quality of formal resources available for particular uses.

: developing students' abilities to draw on resources in accordance with specific performance demands.

: developing students' abilities to interpret features of language in use in target or near-target settings.

: encouraging students to develop their own language learning strategies.

This two strand approach in course design is seen as providing a blend between performance and competence and is reflected in the typical unit structure (POP) in which a unit is divided into lessons :

Lesson 1

Preliminary central activity.

Lesson 2

Focus.

Lesson 3

Exposure.

Lesson 4

Final central activity.

Parts of the unit are seen to have the following functions :

Central activity 1 : to activate students' latent abilities.

: to make students aware of what is involved in a particular type of interaction.

: feedback functions - make students aware of problems in performance, ways of improving it.
- diagnostic value to teachers.

Focus : draw students' attention to certain form/function relationships.

: develops skills of drawing on specific resources for micro-performance and strategies for improving quality and range of resources.

Exposure : develops understanding of language in use (i.e. situation/ topic variations on interaction type) and interpretation of particular features of language in use.

: develops ability to draw on samples of language in use to improve performance.

Central activity 2 : Provides more complex interaction to enable student to draw on resources and strategies built up in support component.

: recycles interaction types previously encountered.

: provide opportunities for teacher/student assessment.[2]

3. Towards Test Design

A face/content valid testing programme would have to mirror the blend between communicative competence and performance as manifested in the courses and focus on use of the language in target or near target settings. Such a programme would need to evaluate actual performance and performance features as well as a display of competence. The testing programme would have to

provide the kind of information required by employers in the professional world. Do the employers want the test results to say that the candidate can perform the required tasks or do they merely want to be told that this person has built up competence and can display this competence in his performance? These two questions have implications for test design and content. None of the three courses - POP, PACE and LAW could lay claims to teaching pure performance. The central activities in the LAW and PACE courses would approximate target events in the professional world. The POP course meant for final year undergraduates would not need to rigourously follow the demands of the professional world. The activities chosen could well be simulations of target events or near approximations of real life situations.

With these considerations in mind it was decided that the testing procedures would include two main approaches - the Operational and Discrete Feature Approach. In the Operational Approach the test items built would be facsimiles of real events in the outside world or profession(s). Task types constructed following this approach would be closest to authentic tasks. The Discrete Feature Approach would test communicative knowledge or competence. The tests would be mainly receptive in nature involving recognition eg. of forms, functions and other enabling skills.

It was agreed that the testing programme for each of the three courses would include the following :

Discrete feature test.

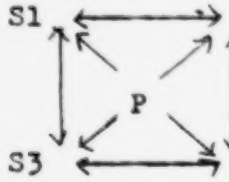
Listening test - tape and answer sheet.

Integrated-skills test - listening-notes-reporting.

Interaction test :

1) Student ⟷ Plant (interlocutor) and Assessor.

2) S1 ⟷ S2 and Assessor



P-Plant (Interlocutor)

4. Test Specifications

4.1 Discrete Feature Test

This test is seen as a tool for formative evaluation and is to be administered after the students have completed a specified number of units in the course. It is a paper and pencil test made up of objective multiple choice items. To ensure content validity the items are modelled on the 'Focus' or support activities in the course units. These items are based on speech utterances (written text) and students are required to choose the most appropriate forms to express the desired functions.

4.2 The Listening Test

The listening test is based on recorded data of either authentic or simulated target events in the

professional world. Listening time is limited to about eight minutes. The item types used are skill-based objective multiple choice items and task-based supply type items.

4.3 The Integrated-Skills Test

In this test the students listen to a taped input of either an authentic or simulated target event. The student is required to take down notes for the purpose of carrying out a specified task e.g. attending a meeting of the Rubber Grower's Council - taking notes of the discussion on 'Housing conditions in the rubber estates' and reporting orally to a superior officer on the main points raised.

4.4 The Interaction Test

The Interaction test is seen as involving :

i) saying something ie. conveying information.

ii) asking something ie. eliciting information.

iii) interacting, which subsumes (i) and (ii) above and involves the testee selecting the right strategy to achieve his aim.

The testing situation would consist of

i) a catalyst or 'plant' - preferably the

teacher who would ease the student into the test and help him along.

ii) an assessor who would not play any part in the interaction.

iii) a student or students (four at the most).

It was felt that the interaction could be between.

i) the teacher or 'catalyst' and the student involving role play.

ii) between two students with the teacher as 'catalyst' and providing help if needed.

iii) between a group of four students with the 'catalyst' providing help.[3]

5. Criteria for Assessing Oral Interaction and Rating Scale.

A study of the specification of criteria of performance (B. J. Carroll 1978) and rating scales currently in use - FSI Rating Procedure, ELTS M3 Interview-Assessment Scale, Bartz Scale (Bartz, 1974), The Schulz Communicative Scale (Schulz, 1977), The Clark Four-Scale System (Clark, 1972) pointed to the need for a less ambitious set of performance criteria and a more modest rating scale.

It was agreed that oral interaction should be assessed on the following criteria - Fluency, Effectiveness, Accuracy and Range. These four criteria were defined as follows : (POP Course).

a) FLUENCY : Smoothness of delivery without obvious groping for words. Continuous and natural speech. Does not refer to speed of delivery.

b) EFFECTIVENESS : Ability to convey meaning and make oneself understood without causing stress on the listener. Ability to employ appropriate turn of phrase and precise vocabulary.

c) ACCURACY : Ability to produce grammatically correct utterances. Pronunciation does not prevent understanding of what is said by any good speaker of the language. Ability to interpret correctly structural and lexical range.

d) RANGE : The ability to draw upon a wide variety of discourse skills, structures as well as lexical items.

A closer look at the definitions shows that the first two criteria - Fluency and Effectiveness are related to the assessment of performance. Accuracy and Range hinge on competence. This is in keeping with the philosophy of course design - the teaching of aspects of competence to enhance performance.

The candidates performance on the test could be judged from two levels - whether he was performing at the Criterion or Target level. The Criterion level is thought of as the level of acceptable performance in terms of course aims and professional or target settings. A candidate whose performance is above criterion level could be judged as performing at different points on the target level - below, about, or at target level. Descriptions of criterion and target level performance were written for each of the four criteria. (Appendix A)

5.1 The Rating Scale

The candidates performance is graded on a scale 1-6. Point 3 equals the criterion, i.e. the level of acceptable performance. Marks of 4, 5 or 6 are awarded to a candidate whose performance is above criterion level. Marks of 1 or 2 are awarded for candidates whose performance is below criterion level. The award of a mark out of 6 is based on a global view of the candidates performance and not on the basis of individual marks awarded for each of the criteria. In tests in which more than one

candidate takes part each candidate is assessed individually and is not penalised or credited for the performance of his/her partner or partners. The scale for grading is matched to specifications of target and criterion level and the Award System in the following manner :

		Grade on Certificate	
Good	6 Target Level	Class I	A
Competent	5	Class II 1	B
Modest	4	Class II 2	C
Marginal	3 Criterion Level	Pass 3	D
Extremely limited	2		E
Intermittent	1		E

6. Test Construction and Trialling

An Oral Interview test and a Group Oral Interaction test were constructed. The tests discussed here are for the POP (Pre-Occupational Purposes) Course meant for final year undergraduates in the Faculty of Arts and Social Sciences and the Faculty of Economics and Administration.

The prime concern in the construction of both these test instruments was the need to reduce the open-endedness of the candidate's reponse. The other consideration was that the candidate should not be penalised because of a paucity of knowledge on the topic under discussion. It

was felt that the provision of a 'brief' for the candidate would serve both ends. The provision of a brief on the topic would set the framework within which he could operate. But within this limited framework he still had the independence to display his competence and performance. It was envisaged that the plant or catalyst would provide him with opportunities to do this.

6.1 The Interview Test

The task set in this oral interaction test was aimed at testing the student's ability to ask for information. The student was told that he had been awarded a scholarship to study for a post-graduate diploma in an English speaking country. He had arranged for a meeting with the students' officer to get some information. The student was given a brief containing a list of topics on which he could ask for information.

The test was trialled on students of diverse levels of proficiency and recordings were made. The trialling showed that the test which concentrated on 'asking for information' led to the plant or informant doing most of the talking. It was felt that the test could be improved by setting questions for the plant to ask the student.

A further observation was that when students have a list of topics on which to ask questions, there is a tendency for the weaker student to say "and what about ....". In the face of this an alternative testing procedure was tried out. Students were given a description of the overall situation and had five minutes to think about the questions they would ask. This procedure produced some amount of interaction but there was no evidence to suggest that it was more effective than the previous one. As pointed out earlier, in this procedure there was the danger that the unimaginative student could be penalised for not being able to think of things to ask questions about.

As it stood the test was only producing a limited kind of interaction. The scope of the test and the ensuing interaction could be broadened by modifying the situation. The student was told that he had already been offered a scholarship by the Public Services Department.

He was provided with full details of the offer - some of which were more favourable and some less favourable than the second offer by an International Agency. The task was for the student to ask questions in order to elicit the information required to compare the two offers. The plant was to ask the

student questions to ascertain the degree of interest in the second offer and to pressure him to make a decision.

This test was tried out on a group of junior officers in the Ministry of Foreign Affairs and recordings were made. A study of the recordings showed that the test was producing the kind of interaction that was intended. With minor modifications in the brief for students the test was ready for use on the target population.

6.2 Group Oral Interaction Test

6.2.1 Test Description

The first test written was based on the units of the POP Course piloted in April 1982. The target event chosen for the test involved 'small-group discussion'.

The student was presented with the following situation. He is a member of a Committee set up by the Residents' Association to study the growing problem of pollution in his area. He has gathered information on one aspect of pollution in his area and the measures needed to remedy the situation. The interactive task set was for him to present his main findings and proposals to the committee, give clarification when required and argue for his proposals to be accepted.

The plant or catalyst was to play the role of chairman. Each student was provided with a brief on one aspect of pollution ... Student A - Water Pollution, Student B - Air Pollution, Student C - Solid Waste Pollution, Student D - Noise Pollution.

6.2.2 Administration

Before the start of the interaction the plant explained the task to the students and answered their questions. At the end of this students were given five minutes to go through the notes in their brief. The first part of the performance was the presentation of the findings and proposals. The next part was the open discussion in which the students were required to argue for the acceptance of their proposals and seek clarification on those put forward by others. The assessor took no part in the discussion and the grade he awarded was on the basis of a global impression of the student's performance in both the presentation and the subsequent discussion. Recordings were made of student performance on the test.

6.2.3 Discussion

Analysis of the recorded data and discussions with 'plants' and assessors showed that the notes in the student briefs were too technical in nature. There was too much of input in terms of the information provided. A student's performance on the test depended very much on whether he understood

the notes provided in the brief. The original intent was that the notes in the brief should aid him in his performance. In practice it had turned out to be otherwise. Here was a fault in the test design that needed correcting.

In the light of this feedback it was decided that subsequent group oral interaction tests would keep the input in student briefs to a minimum. The topic chosen for the interaction should not also be too demanding in terms of lexis.

6.3.1 Test Description

In the next round of piloting in August 1982 another group oral interaction test was written and tried out on a sample of the target population. The tasks which the students were asked to perform were similar to those they had practised in the pilot course. In this test a situation was presented and the students were required to do the following tasks :

a) State their views on an issue.

b) Listen to the views presented by others.

c) Support and justify their own proposals while attacking those put forward by others.

d) Reach a consensus decision within the group.

The situation presented was a familiar one and the test was based on simulation of a small-

group discussion. For actual test instrument see Appendix B.

6.3.2 Administration

The test was administered to groups of four students at a time. The plant or catalyst was to play the role of chairman and coordinate the discussion. The first step in the administration was for the plant to explain the task clearly to students in the group and to answer their questions on the notes in the brief. The students were given five minutes to write short notes and organise their thoughts on the issue presented in the test. Each student then presented his views uninterrupted. The maximum time allowed for this activity was five minutes but most of the students did this in less than three minutes. Fifteen minutes was allotted for discussion of the different proposals. Students were required to support and justify their own proposals while attacking those of the others. At the end of the discussion the plant - acting as leader was, to help the group to reach a concensus decision on the issue or problem. The assessors who did not take any part in the interaction awarded a global score on the basis of the student's performance in the presentation and the subsequent discussion bearing in mind the four criteria - fluency, effectiveness, accuracy and range.

6.3.3 Discussion

The mean score for the test was 4 (on the 6 point scale) and the Standard Deviation was 1.37. Eleven of the students had scores above the criterion level (3 - on the rating scale) and five of them were awarded criterion level scores. The performance of three students was rated at the target level and it is interesting to note that only two had scores below the criterion level.

There was general agreement among the assessors and 'plants' that the test generated a fair amount of interaction. The situation presented was life like and students were able to draw upon their linguistic and communicative competence to perform the required tasks.

An attempt was made to find out whether student performance on the group interaction test would correlate with performance on the Listening test. The coefficient of correlation derived was 0.62. Though modest it gives hope for the future.

7. The Training of Assessors and 'Plants'

Since the reliability of the rating procedure has implications for test validity it was felt necessary to institute a training programme to enhance intra-rater reliability and inter-rater reliability (Clark, 1978).

The decision to use two separate individuals - one for assessing and the other for aiding the

interaction was a first step in this direction. The seperation of functions would not in itself bring about improved reliability in the rating of student performance. Assessors and plants had to be trained in their seperate functions.

7.1 The Training of 'Plants'

The training of plants or catalysts involved first of all a definition of the role. The plant or catalyst would be someone familiar to the students being tested, preferably the teacher. One of his main functions is to put the testee at ease and provide help when it is required. He assumes the role assigned to him in the interaction and gives the testee opportunities to display his competence and performance. He could accomplish this either, by asking probing questions or by being purposefully evasive in withholding information. He is also seen as the instrument for introducing unpredictability into the interaction to see how the testee copes with the situation.

Before the tests were trialled 'would be' plants - teachers who had taught the pilot courses were exposed to recordings of oral interaction tests. They were asked to evaluate the performance of the plant in the test. It was found that some plants tended to dominate the interaction while others provided the testees with the desired opportunities

for performance. In this manner plants were made aware of their role and function. They then simulated the role of the plant in a group oral interaction test.

7.2 The Training of Assessors

Before the Oral Interaction tests were administered assessors were brought together. They were provided with the six-point rating scale and descriptions of the criteria on which the oral interactions were to be assessed - fluency, accuracy, range and effectiveness. They went through the criterion and target level descriptions of the criteria and doubts were ironed out. Assessors were told that they were required to make a global assessment and not to award seperate scores for each criteria.

Next, the assessors listened to recordings of student performance in a group oral interaction test (April 1982 Piloting). Assessors were asked to rate the performance of one selected student bearing in mind the criterion and target level descriptions of the criteria. The scores awarded for this candidate were as follows : 3, 3, 3, 3, 3, 5. The majority of the assessors had awarded this candidate a score of 3 - a criterion level score on the 6 point scale. In the discussion that followed it became apparent that the assessor who

awarded a score of 5 recognised the student's voice and awarded the score on the basis of his performance in class. This assessor agreed that she had been over generous in her assessment.

The scores awarded for the second candidate who was assessed wereas follows : 4, 4, 4, 4, 4, 3. There was general agreement among the assessors that this candidate's performance was above that of criterion level and that he deserved a score of 4 on the 6 point scale. From the trial grading procedure it was evident that assessors had a fairly good grasp of what is expected at criterion level performance on the four criteria.

8. Summary

What has been described in the previous sections is an attempt at ensuring enough validity and enough reliability for the test procedures used in the Project. The decisions made in relation to test validity and reliability are purely judgemental ones and in this there has been a break with tradition.

A concerted attempt has been made to ensure that course goals and aims are reflected in the testing procedures. The situations presented in the oral interaction tests are simulations of target events in the professional world or real life settings. The tasks set for the student demand the use of the language.

The specification of criterion and target levels of performance have been derived from not only a careful study of the principal teaching items, but student data, target occupational operations and other assessment scales. The criteria on which the performance is assessed closely reflects the blend between communicative competence and performance as seen in the materials.

The attempt at improving the reliability of the testing procedure has taken two directions - the first is the refinement of the test instrument and next is the training of assessors and plants . Improvements have been made to test design and format in the light of feedback from student recordings and test administrators observations. Valuable insight has been obtained on the kind of student brief that will allow for freedom while restricting the opendedness of the response. There is evidence that the training of assessors has produced consistency in the rating of student performance. The training of plants is aimed at reducing variance in the use of elicitation techniques.

The practicality of administering this performance testing procedure on a large population has still to be resolved and begs discussion.

It is hoped that the path taken is the right one though beset with puddles!

NOTES

1. I am grateful to John Moore and the staff of the University of Malaya Spoken English Project for the use of Project materials and documents.

2. Most of this section is based on John Moore's paper, "POP Syllabus Rationale", UMSEP/SY/POP/6.

3. The test format discussed in this section was suggested by Robin Davis in his Consultancy Report, UMSEP/CR/5, University of Malaya Spoken English Project.

4. In the context of the University of Malaya Spoken English Project POP refers to the course designed for Pre-Occupational Purposes and PACE refers to the Public Administration and Commercial English course.

5. The term 'Plant' is used in this paper to refer to the interlocutor or catalyst who assists the students in the group oral interaction test.

REFERENCES

Canale, M. (1981). "Communication : How to Evaluate It", Paper presented at 12th Annual Symposium, Carleton University, ottawa, Canada.

Carroll, B. J. (1980). Testing Communicative Performance. Oxford : Pergamon.

Clark, J. L. D. (1972). Foreign Language Testing : Theory and Practice. Philadelphia, Pa : Center for Curriculum Development.

Clark, J. L. D. (1978). 'Direct vs Semi-Direct Tests of Speaking Ability'. In Concepts in Language Testing : Some Recent Studies, edited by Eugene J. Briere and Frances Butler Hinofotis, pp. 35-49. Washington, D.C. : TESOL.

Davis, A. (1972). Language Testing - Survey Article in two parts. Language Teaching and Linguistic Abstracts, Vol. II.

Jones, R. L. and Spolsky, B. (eds) (1975) Testing Language Proficiency. Washington, D.C. : Center for Applied Linguistics.

Oller, J. W. (1979) Language Tests at School. London, Longman.

Valette, R. M. (1977). Modern Language Testing (2nd Edition) Harcourt, Brace, Jovanovich.

END

U.S. Dept. of Education

Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 29, 1991